

# 統計分析のための変数処理ツールの試作

青 木 博 明

## Trial Production of a Processing Variable Tool for Statistical Analysis

Hiroaki AOKI

### Abstract

We will introduce a tool that we have made for processing variables for statistical analysis. We can use this tool by mouse clicking on a Graphic User Interface or by inputting parameters. This tool can be used to calculate the statistics or draw the graphs. It has the advantage of proceeding data and explanation for variables from the very beginning. We hope that it will be used to assist students learning statistics and other subjects, especially economics and business management.

### 1. は じ め に

本稿では筆者が作成した変数処理用ツールの紹介を行う。このツールは統計分析を前提とするものであり、その特徴は、まずGUI（グラフィック・ユーザー・インターフェース）での操作を念頭においている点にある。したがって従来型のCUI（キャラクタ・ユーザー・インターフェース/コマンドライン・インターフェイス）による統計ソフトとは異なり、命令文や計算式の入力を必要とせず、表示・計算のコマンド・ボタンやオプションへのマウスクリックとパラメータの入力だけで操作が可能である。具体的な機能はデータの選択、数値表示、グラフ表示、簡単な統計分析である。なお本ツールはVisual Basic (ver6.0) で作成している。

従来型のCUIによる統計ソフトやExcelなどの統計処理機能を持つソフトの場合、命令の文法、関数や計算式の機能や書式を習得しなければいけなく、それに労力と時間を費やす必要がある。本ツールではそれをかなり削減できるものとする。ただしそれら命令文や関数、計算式によって計算分析する統計処理可能なソフトに較べると、ユーザーが独自に自由に詳細な分析手順を組むのには適さない。

加えて本ツールは、データを全て変数単位で処理するという特徴を持つ。変数単位でデータを読み込み、それに続くデータの表示、加工、分析も全て変数単位で行う。また本ツールでは、変数の集合を変数群と呼び、この変数群をより上位の処理単位としている。

さらに、ツール自体が最初から変数もしくは変数群をデータとして持つというのが大きな特徴である。また重要な変数についてはその変数に関する短い説明が表示される。変数は筆者が選んだ経済関連の変数で、ユーザーはそのデータの取得に時間を費やす必要がなくなる。このようにツールが最初から変数を持っていることは当然ながら便利なことであり、この点はこのツールが画面上のクリック一つで、変数の読み込み、表示・計算ができるという操作の利便性と相乗効果を持つものといえる。また異なる変数群への切り替えが可能で、これによってさらに多彩な変数を読み込むことができる。

そしてそれら変数の内から分析加工のための変数を選択して、自分用の変数群に登録し、ファイルとして保存呼出することができる。

なおユーザー自身が独自のプライベートな変数群もしくは変数を追加することもできる。ただしその場合はExcelによるデータ入力が必要である（これに関する詳しい説明は割愛する）。

このようにそれ自体が最初から変数としてのデータを持ち、GUI上で簡単に操作できるツールを作成しようという筆者の動機は、大学での講義の経験から生まれたものである。統計学の講義で、経済関連のデータを扱うとき、その検索や準備に手間がかかるのだが、ツール自体がデータを持っていると、講師及び受講生はそれを用意する煩わしさから逃れられる。また各変数についている説明文が変数の意味を確認させ、知識の習得に役に立つはずである。またソフトの操作自体の習得に時間をとられることが少ない。

現在ツールは経済経営関連の変数群のみをいくつか持つ。それは筆者が経済を教えているからであって、もちろん労を惜しまなければ、経済経営以外の分野の変数を用意することもでき、その分野に関しても同様に自主学習や講義において利用することができる。

本ツールは、現在のところ変数の選択・表示・グラフ化及び基本的な統計処理の機能しか持たない。しかし今後、このツールを基にして、さらに高度な統計分析である重回帰、因子分析、主成分分析、各種の推定・検定の計算機能を付加する予定である。その場合には、このツールが現在持つ変数の表示、選択の方法に準拠する予定である。冒頭で、本ツールは統計分析を前提とする、と書いたのには、その意味も込められている。

以下で、2. ツールの設計上の特徴、3. ツールの操作画面と機能、を説明していく。

## 2. ツールの設計上の特徴

本ツールにおける設計上の特徴と考え方をいくつかの項目に分けて述べる。すでにその概要は述べたので、ここではより具体的な説明をする。

## 2.1 変数群と変数ユニット

本ツールでは処理の対象となるデータを、個々の数値単位ではなく、変数を単位として、表示し処理する。このように個々の数値ではなく、変数による処理上の単位を、ここでは変数ユニットもしくは単にユニットと呼ぶ。

さらにこの変数ユニットを複数その中に含むのが変数群である。例えば変数群として「マクロ経済」があるが、これはマクロ経済に関するGNP、為替レート、CPIなどの変数を含む。その他にも変数群として「企業」「国別」などを用意している。ユーザーはマウスクリックによって、画面上で変数群を選び、さらにその中から必要とする変数を選び、表示・加工・分析の対象とすることになる。重要な変数については変数に関する短い説明文の表示も行われる。

変数群はデータファイルから読み込まれるのだが、このデータファイルはあらかじめExcelで作っている。まず原データをExcelのワークシートに用意して、それをExcel VBAで自動的にテキストファイルに変換し、それを本ツールが読み込む、という手順である。ユーザーは、後で述べるプライベート・データを扱わない限り、この過程に一切かかわらない。ただしツール上で個々のデータを見ることはできる。

### [変数の属性]

変数はそれが持つデータに応じた属性を持つ。例えば、属性の例としては、TY（年次の時系列）、TM（月次の時系列）などがある。他にクロスセクション・データとして、C(国)、E(企業)がある。一般に、元の変数群が異なっても、属性が同じならば変数は同一のユーザー変数群（後出）に入れることができ、同じ分析の対象とすることができるが、属性が異なるとその限りではない。

### [データ欠損値と期間のズレ]

データの欠損や時系列の期間のズレは分析上煩わしい問題であるが、それにも対処できるようにしている。ただし欠損値を補完するわけではなく、データがないところを飛ばして、変数間でデータが共通に存在する範囲に関してのみ計算処理するのである。期間のズレについても、共通にデータが存在する範囲に関してのみ処理するようにしている。

## 2.2 パブリック変数とプライベート変数

本ツールでは変数をパブリック変数とプライベート変数に分ける。パブリック変数とは、一般に公開されており、全ユーザーが共通に閲覧利用できる変数である。例えばマクロ経済データや上場企業のデータ、株価データなどである。もちろん経済関連以外のデータも用意することができる。これらのデータはツールがあらかじめ持っているので、ユーザーが用意する必要はない。

他方、プライベート変数とは、個々のユーザー（人または組織）が占有して閲覧利用する変数である。その利用のためには、それらの変数のデータを、ユーザーがある書式にしたがって

Excel上に用意することが必要になる。もちろん、パブリック変数についてもユーザーがExcel上に用意することで、追加することができる。

なお企業などの組織で、複数のユーザーがプライベート変数を使う場合、だれか一人がその変数を用意すればそれを共有できるわけであるから、例えばある企業の商品別の売上量をツール担当者が用意すれば、企業内の上層部から下位の社員までがともにそれを閲覧利用できるわけである。

### 2.3 基礎統計・グラフの計算表示

ツールの起動時に、各変数の始期データ位置、終期データ位置、平均、最大、最小をツールが計算して、これらを表示し、またその後の計算に利用する。

基本的統計量として、変数ごとに平均・分散・標準偏差を表示し、またグラフを表示する。グラフには特別な機能（後出）をつけている。グラフの表示の有無については、オプションでの選択が可能である。

### 2.4 変数の説明文

重要な変数については、それに関する簡単な説明文が表示される。これは講義や自習学習においても役立つものと考えられる。上述のように変数に説明文が付いていることは、ツールが最初から変数データを持っていることと相乗効果を持つといえる。

### 2.5 ユーザー変数群 (GX GY)

ツールが持つ変数群の内から任意の変数を選んで、ユーザー独自の変数群に登録することができる。これをユーザー変数群と名付ける。これに対して、最初からツールが持つ変数群を一般変数群と呼ぶ。ただし、上述のように分析上の理由から、同じ属性の変数でなければ、同じユーザー群に入れることができない。ユーザー変数群にはGXとGYがあり、さらにこれらに名前をつけてファイルとして保存呼出ができる。よってGXとGYとも複数作ることができる。GXが基本であり、必要ならばGXの中からGYを選ぶようになっている。ただしGYは一つの変数しか登録できない。(実際には現在のところGYの変数はツールの計算処理の対象とはなっていない。ただし今後予定しているツールの機能拡張によって、例えば回帰分析の説明変数などにする予定であり、それを示すためにあえて本稿で表示した。)

ユーザー群GXに入れた変数に関しては、Log（自然対数）、ベキ数（指数）、タイムラグによる加工が可能である。GXに登録し加工した変数がその後の統計分析の対象となる。

### 3. ツールの操作画面と機能

ツールの各画面とそれに対応する機能・操作方法を並行して説明していく。

以下で、画面上のリストは変数の一覧、BOXはデータ・グラフ・テキストなどの表示を行う。F2.1~F2.5はコマンドで、それをマウスクリックすると対応する機能を実行する。その他画面上のオブジェクトとしてオプション、チェックがある。基本的に（マウス）クリックとパラメータの入力で操作を進める。

図1 A 変数選択画面 「説明文・ユーザー群」

**変数選択**    計算分析    変数群    ツールの説明

表示オプション  
 説明文  
 ユーザー群  
 グラフ  
 データ

グラフラベル チェック  
 テーマ  
 項目

変数群=マクロ経済

| 変数名   | 属性 | データ数 | 平均   | 範囲        |
|-------|----|------|------|-----------|
| 時系列1  | TY | 20   | 3.05 | 1982~2001 |
| GDP   | TY | 20   | 3.05 | 1982~2001 |
| GNP   | TY | 16   | 3.81 | 1984~1999 |
| 輸出総額  | TY | 16   | 3.81 | 1984~1999 |
| 輸入総額  | TY | 16   | 3.81 | 1984~1999 |
| 為替レート | TY | 10   | 6.10 | 1986~1995 |
| 消費者物価 | TY | 10   | 6.10 | 1982~1991 |
| 卸売物価  | TY | 16   | 3.81 | 1982~1997 |
| 生産指数  | TY | 10   | 6.10 | 1982~1991 |
| 景気指数  | TY | 10   | 6.10 | 1982~1991 |
| 就業者   | TY | 20   | 3.05 | 1982~2001 |
| 完全失業者 | TY | 20   | 3.05 | 1982~2001 |
| 完全失業率 | TY | 20   | 3.05 | 1982~2001 |

変数名 = GDP    属性 = TY    単位 = 兆円    データ数 = 20  
 範囲 = 1982~2001    最小 = 274.7078    最大 = 523.0511  
 平均値 = 3.05    標準偏差 = 437.222

< 帰属計算 >    [変数の基本データBOX]  
 定数項 A = 2.39    傾き B = .22    決定係数 = .96    相関係数 = .98

国内総生産。一年間である国の国内で生み出された付加価値の総額。GNP(国民総生産)から海外純所得を引いた金額でもある。国の経済活動の指標として用いられる。  
 [変数説明文BOX]

ユーザー変数群

X 輸出総額

⇒ GDP  
 GNP  
 輸出総額  
 輸入総額

Del

Y指定

変数の数 = 4    変数属性 = TY

GX 変数変換

ベキ数    変数  
 log    1    2

ユーザー変数群 GY

⇒ GDP

Del

ユーザー変数群 ファイル処理

ファイル読み込み    ファイル書き込み

図1B 変数選択画面 「グラフ・データ」

Form1

変数選択    計算分析    変数群    ツールの説明

表示オプション  
 説明文  
 ユーザー群  
 グラフ  
 データ  
 項目

変数リスト    変数群=マクロ経済

| 変数名   | 属性 | データ数 | 平均   | 範囲        |
|-------|----|------|------|-----------|
| 時系列1  | TY | 20   | 3.05 | 1982~2001 |
| GDP   | TY | 20   | 3.05 | 1982~2001 |
| GNP   | TY | 16   | 3.81 | 1984~1999 |
| 輸出総額  | TY | 16   | 3.81 | 1984~1999 |
| 輸入総額  | TY | 16   | 3.81 | 1984~1999 |
| 為替レート | TY | 10   | 6.10 | 1986~1995 |
| 消費者物価 | TY | 10   | 6.10 | 1982~1991 |
| 卸売物価  | TY | 16   | 3.81 | 1982~1997 |

変数名=GDP    属性=TY    単位=兆円    データ数=20  
 範囲=1982~2001    最小=274.7078    最大=523.0511  
 平均値=3.05    標準偏差=437.222

< 回帰計算 >    [変数の基本データBOX]  
 定数項 A=2.39    傾き B=.22    決定係数=.96    相関係数=.96

GDP    [グラフBOX]

Max=523.0511

1986    341.89

1990    442.07

1982    2001

データBOX (選択変数のデータ)

|      | GDP   | 偏差値  | 指数(初年) |
|------|-------|------|--------|
| 1982 | 274.7 | 56.2 | 100.0  |
| 1983 | 286.5 | 56.5 | 104.3  |
| 1984 | 305.8 | 56.9 | 111.3  |
| 1985 | 326.8 | 57.4 | 119.0  |
| 1986 | 341.9 | 57.7 | 124.5  |
| 1987 | 356.5 | 58.1 | 129.8  |
| 1988 | 382.5 | 58.7 | 139.2  |
| 1989 | 410.5 | 59.3 | 149.4  |
| 1990 | 442.1 | 60.0 | 160.9  |
| 1991 | 469.9 | 60.7 | 171.1  |
| 1992 | 482.4 | 61.0 | 175.6  |
| 1993 | 486.4 | 61.1 | 177.1  |
| 1994 | 491.9 | 61.2 | 179.1  |
| 1995 | 498.9 | 61.3 | 181.0  |
| 1996 | 511.9 | 61.6 | 186.4  |
| 1997 | 523.1 | 61.9 | 190.4  |
| 1998 | 516.6 | 61.7 | 188.0  |
| 1999 | 509.1 | 61.6 | 185.3  |
| 2000 | 513.4 | 61.7 | 186.9  |
| 2001 | 507.5 | 61.5 | 184.7  |

<F1> [変数選択 画面]    図1A、図1Bを参照

図1Aもしくは図1Bによって例示される画面が最初に立ち上がる。この画面において表示される変数の一覧からある変数をクリックで選択すると、その変数に関する属性、数値データ、基本統計量、グラフが表示される。またユーザー群への登録が可能で、そのユーザー群をユーザーはファイルとして保存することができる。図1Aと図1Bの画面は、後述する[表示オプション]によって切り替わる。

以下、この画面における各コントロールとその機能を説明する。

#### F1.1 [変数リスト] <表示された変数群からの変数選択>

[変数リスト]に、指定された変数群に登録されている変数が一覧される(図では変数群「マクロ経済」を表示している)。変数リストには、変数の変数名、属性、データ数、平均値、

(データの存在) 範囲が表示される。この変数リスト上である変数をクリックすると、次に示すように [変数の基本データBOX] に各種のデータが表示される。

#### F 1.2 [変数の基本データ BOX] ([変数リスト] のすぐ下)

[変数リスト] にある変数を一つクリックすると、ここに変数名、属性、存在するデータ数、(データの存在) 範囲、最大値、最小値と基本統計量である平均、標準偏差が表示される。

また時系列データの場合は、時系列を説明変数、クリックした変数を非説明変数とする回帰直線の定数項、係数、決定係数が表示される。

#### F 1.3 [表示オプション]

この [表示オプション] で、「説明文・ユーザー群」を選ぶと、選択した変数に関する短い説明文を表示する [変数説明文 BOX] が画面下に、ユーザー群のGXとGYの変数リストが画面右に表示され、他方「グラフ・データ」を選ぶと、選択した変数に関するグラフ (横軸に時系列など、縦軸にデータの数値、後出) を描画する [グラフ BOX] が画面下に、その数値が画面右の [データBOX] に表示される。数値は、原データと偏差値、それとデータの最初の値を100として標準化した指数が表示される。初期のオプションは「説明文・ユーザー群」である。

図1 Aが「説明文・ユーザー群」を選んだときの画面で、図1 Bが「グラフ・データ」を選んだときの画面である。

#### F 1.4 [変数説明文 BOX] (図1 A [変数の基本データBOX] のすぐ下)

[表示オプション] で「説明文・ユーザー群」を選ぶと、[変数リスト] で選んだ変数に対する簡単な説明文が表示される (説明文が用意されていない場合もある)。これはツールが変数のデータとともに最初から備えているものである。

#### F 1.5 [グラフ BOX] (図1 B [変数の基本データBOX] のすぐ下)

[表示オプション] で「グラフ・データ」を選ぶと、[グラフBOX] と後述の [データBOX] が表示される。[グラフBOX] には選択した変数に対する点グラフと数値が表示される。点グラフといっても実際には小さな円が表示されるのだが、この小さな円のどれか一つをクリックすると、その円に対応する [データBOX] 上の項目に赤い下線が引かれる。

このグラフの座標はデータの最大値によって、標準化しており、データの大きさにかかわらず、画面におさまるようにしている。

### F 1.6 [グラフラベル チェック]

[グラフBOX]におけるグラフのグラフラベルの表示に関する選択を行う。グラフラベルとして選べるのはデータと項目である。データをチェックすると、グラフの小さな円の横にデータが表示され、項目をチェックすると項目名（例えば年次）が表示される。二つ同時の表示も可能である。

### F 1.7 [データ BOX] (選択変数のデータ)

[表示オプション]で「グラフ・データ」を選ぶと、後出の[ユーザー変数群]に代わって、このBOXが向かって右側に表示される。表示される数値は、原データと偏差値、それとデータの最初の値を100として標準化した指数である。

### F 1.8 [ユーザー変数群] < [ユーザー変数群 GX] [ユーザー変数群 GY] >

[表示オプション]で「説明・ユーザー群」を選ぶと、[データBOX]に代わって、これが画面右側に表示される。[変数リスト]上である変数を選んだ状態で、「⇒」ボタンをクリックすると、その変数が画面右側のユーザー変数群GXにコピーされ登録される（ただし、本稿の図では、[ユーザー変数群GX]の表示の代わりに、GXで選択された変数名が表示されている）。このGXは複数の変数を含むことができ、それらの変数を対象に、次の変数分析で示す各種の統計分析を行うことができる。なお、最初に選ばれたものと同じ属性の変数でなければ、登録を拒否される。GXのすぐ下に登録された変数の数と属性が表示される。

「Del」をクリックするとGXからその変数が削除される。「Y指定」をクリックすると変数がGXからさらにユーザー変数群GYにコピーされる。やはり「Del」によってGYの登録をはずすことができる。GYには一つの変数しか登録することができない。（上述のようにGYへの登録は現在どの統計処理にも対応しない）

#### F 1.8.1 [ユーザー群 変数変換]

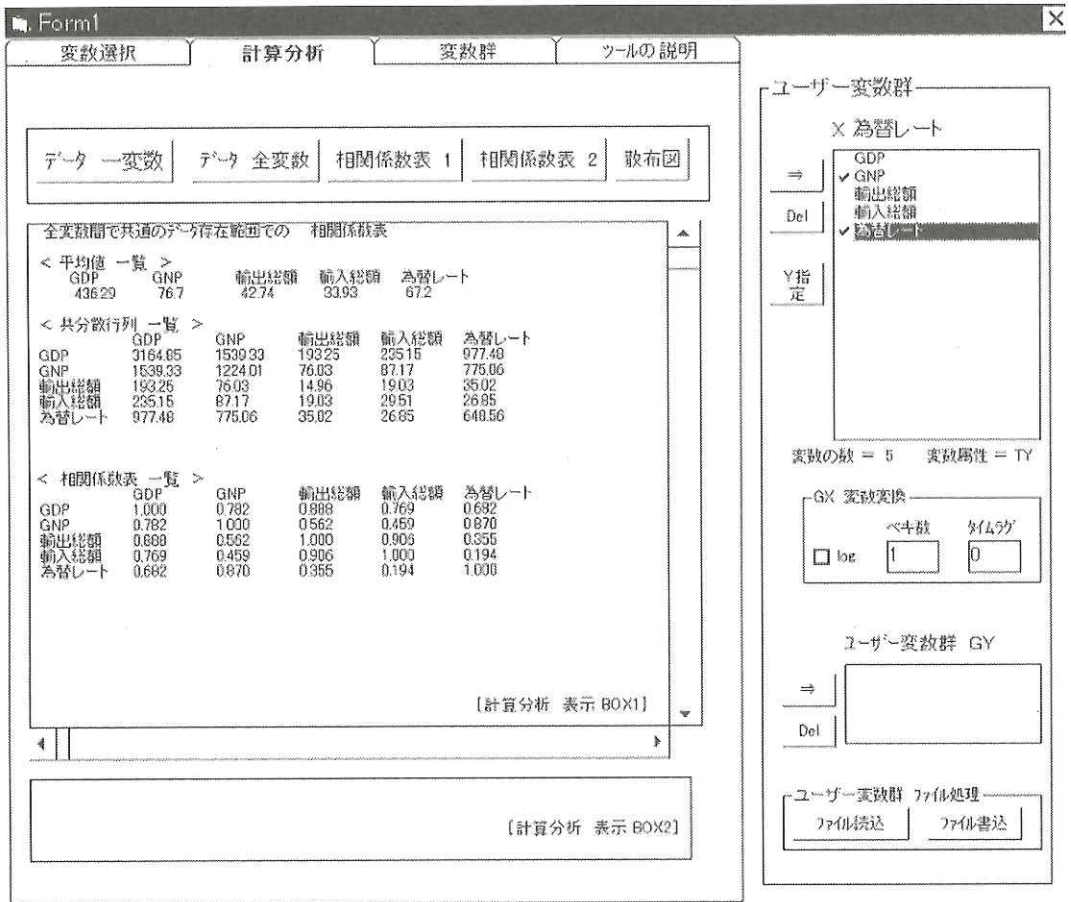
GXに登録された変数それぞれに対して、Log（自然対数）、ベキ数（指数）、タイムラグによる加工変換が可能である。これらの加工変換は回帰分析などにおいてよく行われるものである。加工変換後の数値がツールでの計算の対処となる。加工変換はチェックや整数の代入によって行う。

#### F 1.8.2 [ユーザー変数群ファイル処理] <GX GY群の保存呼出>

ユーザー変数群GX、GYをファイルとして保存呼出する。



図2 計算分析 画面 [相関係数表1]



<F2> [計算分析 画面] 図2を参照

ユーザー変数群のGXに登録され加工された変数に対して、原データと加工データ、平均値、共分散表、相関係数表、散布図の表示と計算を行う。相関係数表には2種類あり、計算の対象とする範囲が異なっている。図2では[相関係数表1]の例を示している。以下詳しい説明を行う。

### F.2.1 [データ 一変数] コマンド

このボタンをクリックすると、GXのリストで選択された変数の原データ (C) と加工後 (T) のデータを表示する。画面のスクロールが可能である。

### F.2.2 [データ 全変数] コマンド

GXに登録された全変数の原データと加工後のデータを表示する。画面のスクロールが可能である。

### F 2.3 [相関係数表 1] コマンド

GXに登録された全変数の加工後のデータに関する平均値と共分散表、相関係数表を計算し表示する。共分散表と相関係数表は全変数間でデータが共通に存在する範囲で計算している。

### F 2.4 [相関係数表 2] コマンド

上の相関係数表 1 と同様、GXに登録された全変数の加工後のデータに関する相関係数表を表示する。ただし [相関係数表 1] と違って、対象となる 2 変数間でデータが共通に存在する範囲で計算を行っている。

### F 2.5 [散布図] コマンド

GXに登録されている変数の内の二つにチェックを入れると、その二つの変数に関して、散布図を表示し、また相関係数の計算を行う。

### F 2.6 [計算分析 表示 BOX 1]

上記の各種コマンドに応じた数値・グラフを表示する。全体を見るために、スクロールバーによる画面のスクロールが可能である。

### F 2.7 [計算分析 表示BOX 2]

[計算分析 表示BOX 1] に付随する情報を表示する。

#### <F3> [変数群 画面] 図は省略

ここでは複数の変数群が表示されるので、その中から一つを選ぶと、その変数群が持つ変数を [変数リスト] に表示する。今のところ、ツール起動時に表示される変数群は「マクロ経済」に固定している。

#### <F4> [ツール説明画面] 図は省略

ツールの操作と機能を説明する文章を表示する。ツール説明用のドキュメント・ファイルにリンクしている。テキスト画面をダブルクリックするとワードが立ち上がり、ファイル全体を見ることができる。

その他の注意点を述べる。本稿の画面で示した変数や変数群は仮のものであり、また数が少ない。紙面では、ツールの操作画面は白黒であるが、実際にはカラーである。本稿の白黒の紙面印刷用に、本来のツールの画面を少し変更している部分がある。

#### 4 む す び に

最初に述べたように、今後このツールを基にして、さらに高度な統計分析である重回帰、因子分析、主成分分析、各種の推定・検定の計算機能を付加する予定である。その際も、このツールが持つ変数の表示、選択の方法に準拠する予定である。

その他、新機能の拡張というよりは、現段階でツールを変更改善できる点として、具体的に次のようなものを考えている。

グラフについては、Excelに無い機能も付加しているが、さらに有用で操作性に優れた機能を付けることが可能であろう。整然としたデータ表示には技術的に難しいところがある。現時点でもツール画面の変数リストに若干の不揃いがある。現段階では、データ表示はPictureコントロールへの書き込みで行っているが、表形式での表示やExcelとのデータの互換性、ファイル管理の簡便性を考えるとMSFlexGridコントロールの利用も有効であると考えている。ある変数の増加率の表示機能や、異常値の明示や削除機能も検討中である。画面デザインの改善も必要だと考える。

ツールは、操作が簡便で起動や動きが軽いことが、一つの重要な要素である。その点を念頭におきつつ、さらなるツールの改善を目指すつもりである。

このツールは、ツール自体が変数のデータや説明を持つことから、変数や変数群を追加すれば、統計学や経済学以外の講義や自習においても、有効に活用できるものと思っている。また、近々ツールが持つ変数を増やすつもりでいる。プライベート変数の利用例についても具体的に説明しようと思ったが、残念ながらそれはかなわなかった。

以上のような改善、変数の充実によって、このツールをより実用的なものにしたいと思っている。

#### 参 考 文 献

- 田中豊, 垂水共之『統計解析ハンドブック 多変量解析』共立出版, 1995.
- 林晴比古『新Visual Basic入門 ビギナー編 Ver.6.0対応』ソフトバンクパブリッシング, 1998.
- 林晴比古『新Visual Basic入門 シニア編 Ver.6.0対応』ソフトバンクパブリッシング, 1998.